



Center for
Educator Compensation Reform

**THE OTHER 69 PERCENT:
FAIRLY REWARDING THE PERFORMANCE
OF TEACHERS OF NON-TESTED SUBJECTS
AND GRADES**

Cynthia D. Prince, Vanderbilt University

Patrick J. Schuermann, Vanderbilt University

James W. Guthrie, Vanderbilt University

Peter J. Witham, Vanderbilt University

Anthony T. Milanowski, University of Wisconsin-Madison

Christopher A. Thorn, University of Wisconsin-Madison

The Center for Educator Compensation Reform (CECR) was awarded to Westat – in partnership with Learning Point Associates, Synergy Enterprises Inc., Vanderbilt University, and the University of Wisconsin – by the U.S. Department of Education in October 2006.

The primary purpose of CECR is to support the Teacher Incentive Fund grantees with their implementation efforts through the provision of ongoing technical assistance and the development and dissemination of timely resources. CECR also is charged with raising national awareness of alternative and effective strategies for educator compensation through a newsletter, a Web-based clearinghouse, and other outreach activities.

This work was originally produced in whole or in part by the CECR with funds from the U.S. Department of Education under contract number ED-06-CO-0110. The content does not necessarily reflect the position or policy of CECR or the Department of Education, nor does mention or visual representation of trade names, commercial products, or organizations imply endorsement by CECR or the federal government.

This report is in the public domain. Authorization to reproduce it in whole or in part is granted. While permission to reprint this publication is not necessary, the suggested citation is: Prince, C.D., Schuermann, P.J., Guthrie, J.W., Witham, P.J., Milanowski, A.T., & Thorn, C.A., *The Other 69 Percent: Fairly Rewarding the Performance of Teachers of Non-Tested Subjects and Grades*. Center for Educator Compensation Reform. U.S. Department of Education, Office of Elementary and Secondary Education, Washington, D.C., 2008.

THE OTHER 69 PERCENT: FAIRLY REWARDING THE PERFORMANCE OF TEACHERS OF NON-TESTED SUBJECTS AND GRADES

Executive Summary

This module addresses the means by which states, districts, and schools can fairly and effectively measure the performance of teachers who teach a grade level or subject area in which standardized achievement tests are not administered or who teach English language learners or students with disabilities.

Most education performance-pay plans rely to a large extent on student scores on standardized achievement tests to identify teachers who are highly effective at increasing student learning. These tests primarily measure student reading and mathematics achievement, although a smaller number also measure additional core subjects such as science and social studies.

Identifying highly effective teachers of subjects, grades, and students who are not tested with standardized achievement tests—such as teachers of art, music, physical education, foreign languages, pre-K to Grade 2, high school, English language learners, and students with disabilities—necessitates a different approach. It is important that states and districts provide viable options for measuring the progress of these groups of students and the productivity of their teachers, both of which contribute to school performance.

Identifying a workable and equitable solution is critical in order to design a performance-based pay system that teachers will not immediately dismiss as unfair. This document provides a variety of specific options for integrating each of these groups of teachers more fully into systems that seek to assess and reward teacher performance.

Potential ways to award performance-based pay vary depending on the type of teacher. Options discussed in this document are as follows.

Teachers of Non-tested Subjects

- Teachers are eligible for schoolwide performance bonuses only.
- Teachers are eligible for some, but not all, of the individual performance bonuses that teachers of core academic subjects are eligible to receive.
- Student test scores are not used to determine non-core teachers' eligibility for rewards. Instead, eligibility is based exclusively on non-test measures, such as observed evaluations of classroom performance, acquisition of additional knowledge and skills, or assumption of additional roles or responsibilities. All of these, and other non-standardized test measures, may be displayed in a thoughtful portfolio.
- New student tests are created to assess teacher performance in non-core subjects.

Teachers of Non-tested Grades: Pre-K to Grade 2

- Create a developmentally appropriate rubric to assess how well teachers in pre-kindergarten to Grade 2 are supporting young children's development on dimensions such as students' cognitive development, social-emotional development, motor development, and language acquisition.
- Use student results from adaptive tests such as the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) or Measures of Academic Progress (MAP) to assess teacher performance at the early grades.
- Use measures other than individual classroom achievement as a way to include teachers of pre-K to Grade 2 in the compensation system.

Teachers of Non-tested Grades: High school

- Use existing tests that were developed for other purposes to estimate teacher contributions to student learning in grades that do not administer standardized achievement tests.
- Adopt or develop new end-of-course tests.
- Base rewards for high school teachers on department-wide performance, rather than individual classroom performance.
- Supplement student test outcomes with schoolwide measures such as high school dropout and graduation rates.

Teachers of English Language Learners

- Base performance rewards for teachers of English language learners on schoolwide achievement gains or reward them by team when the performance of English language learners improves.
- Use student gains in English language proficiency, in addition to gains in subject matter knowledge, as an additional performance measure for teachers of English language learners.
- Use knowledge and skills-based pay structures to reward teachers of English language learners for their expertise.

Teachers of Students with Disabilities

- Base performance rewards for teachers of students with disabilities on schoolwide achievement gains.
- Reward teacher teams when the performance of students with disabilities improves.
- Develop a new “student sharing” average to assess the performance of special education teachers.

This document stresses that each of these strategies has both strengths and weaknesses, and offers the following recommendations to help states, districts, and schools consider how to include the other 69 percent of teachers in performance-pay plans:

- **Consider fairness from multiple perspectives.** All teachers must have an opportunity to earn awards for their contribution to student performance, regardless of the grade, subject, or types of students they teach. However, it is not necessary that access to rewards needs to be equal in every case.
- **Think systemically and holistically.** Provide support and professional development to help teachers meet growth targets and to ensure that the growth targets are linked directly to school and district goals.
- **Include multiple measures and award types.** Consider the advantages of using multiple measures of student performance and teacher effectiveness and the benefits of awards at the individual, team, department, and school levels.
- **Weigh transparency against accuracy.** States and districts will need to make difficult decisions regarding the inclusion of measures that cannot be readily explained to key constituents or that cannot be verified as valid and reliable measures of teacher effectiveness. These trade-offs often hinge on the degree to which measures exhibit

a necessary level of technical rigor while also remaining understandable to those whom they directly affect.

- **Be aware of potential unintended consequences.** In a performance-pay system, the manner in which teacher effectiveness and student performance are measured will influence teacher motivation. States and districts should consider the manner in which the system influences the behavior of teachers at all grade levels, subject areas, and across the spectrum of student ability levels and demographic characteristics.

In all cases, states and districts should consider the degree to which measures used in a performance-pay plan: (1) provide indicators of teacher effectiveness, (2) supply teachers with feedback to encourage their professional development, (3) are part of an integrated system designed to support and develop human capital, and (4) are sufficiently defined and aligned with goals for effective instruction and student learning across the curriculum.

THE OTHER 69 PERCENT: FAIRLY REWARDING THE PERFORMANCE OF TEACHERS OF NON-TESTED SUBJECTS AND GRADES

Options for including teachers of non-tested subjects and grades, English language learners, and students with disabilities in performance-pay systems

No matter how well designed a new teacher compensation system may be and no matter how much support it has from policymakers and the public, it will not be accepted by teachers if they do not believe that it is fair. One of the criteria that teachers typically use to determine whether a performance-based compensation system is fair is whether all teachers, regardless of grade, subject area, or students taught, have an opportunity to earn bonus pay or rewards.

But views on fairness and how different teachers should be rewarded in an alternative teacher compensation system differ. One point of view is that incentive systems should align teacher rewards to the most important instructional priorities of the school, generally student learning gains in the core subjects.

An alternative point of view holds that incentive systems should be designed to encourage every teacher, not just those who teach core academic subjects, to excel in his or her particular subject. In fact, only a very small percentage of public school teachers would qualify for performance awards if eligibility were restricted only to those who teach the subjects that are assessed on state-mandated achievement tests. During the 2004-05 school year, for example, only 31 percent of Florida classroom teachers taught reading and math.¹ As another example, only 15 percent of the staff who taught in large high schools in Alaska were responsible for teaching reading, writing, and mathematics during 2005-06.²

TeacherSolutions, a team of 18 highly accomplished teachers who support performance-based pay for teachers, argues that:

“Singling out only a small proportion of educators for special rewards will never produce the large workforce we need to staff every public school with high-quality teachers . . . Don’t limit rewards only to teachers who teach tested subjects, such as reading and math. If we want excellence across the entire school, we need to create incentive systems that encourage every teacher in every subject to excel.”³

But how can states and school districts measure the effectiveness of the vast majority of teachers who teach non-tested subjects or who teach grades in which standardized achievement tests are not administered?

Most performance-pay plans rely to a large extent on student scores on standardized achievement tests to identify teachers who are highly effective at increasing student learning. These tests primarily measure student reading and mathematics achievement, although a smaller number also measure additional core subjects such as science and social studies. Identifying highly effective teachers of subjects that are not tested with standardized achievement tests—such as teachers of art, music, physical education, vocational education, and foreign languages—requires a different approach; however, finding a workable approach is critical in order to design a pay system that teachers will not immediately dismiss as unfair.

Related and equally important challenges are:

- assessing the performance of teachers in grades that are not usually tested (particularly in pre-kindergarten to Grade 2 and in high school); and
- assessing the performance of teachers of English language learners (ELLs) and teachers of students with disabilities.

Measuring the productivity of these groups of teachers presents enormous technical and political challenges.

Although a number of researchers and education policy analysts acknowledge that measuring the productivity of these groups of teachers presents enormous technical and political challenges, few have offered workable solutions. The U.S. Department of Education and the Center for Educator Compensation Reform (CECR) addressed this need by convening an advisory group of researchers and practitioners with expertise in teacher compensation, performance measurement, testing, education policy, special education, and the education of English language learners.ⁱ

The charge to the assembled group was to recommend ways in which states and local school districts might develop fair and effective measures of performance for teachers of non-tested subjects and grades, as well as teachers of English language learners and students with disabilities. This module draws from the ideas, recommendations, and caveats discussed by the group when they convened in Washington, DC, in May 2008. It also highlights information presented by staff from three school districts that participated in a panel discussion on this issue at the June 2008 meeting of Teacher Incentive Fund grantees.

Readers should note that teacher opinion on this subject is far from uniform. As the comments included in this document will show, what one

ⁱ The members of the CECR advisory group on teachers of non-tested subjects and grades and teachers of English language learners and students with disabilities are: Chris Barr, Columbus (OH) City Schools; Andrej Beijulin, Eagle County (CO) Public Schools; James DiPerna, Pennsylvania State University; Lynn Malarz, National Education Association; Howard Nelson, American Federation of Teachers; Chidi Onyia, Lynwood (CA) Unified School District; Gary Ritter, University of Arkansas; Laura Snyder, North Carolina Department of Education; and Martha Thurlow, University of Minnesota.

teacher believes to be a fair way to recognize and reward effective teachers may be considered unfair by others. Valid tensions exist between compensation systems that treat all teachers alike and those that afford teachers greater access to rewards if they perform jobs considered more central to the strategy of improving school performance or if they achieve outcomes that are more highly valued.

The opinions of other experts on this subject also vary widely, and all the advisory group members may not endorse all strategies discussed in this paper. Some of the advisory group members explicitly stated during the meeting that they did not endorse a particular strategy, but were simply raising it for group discussion. All of the strategies described in this paper have both strengths and weaknesses; none is an ideal solution. However, the ideas presented and the lessons learned from these early efforts are certainly worth considering when attempting to design a performance-pay program that includes all teachers.

Organization of this Module

This module is organized in four sections to describe the challenges entailed in measuring the performance of teachers who teach:

1. non-tested subjects (e.g., art, music, physical education);
2. non-tested grades (e.g., pre-kindergarten to Grade 2 and high school);
3. English language learners; and
4. students with disabilities.

Each section explains:

1. why it is difficult to assess the effectiveness of this particular group of teachers;
2. potential ways to measure their productivity; and
3. advantages and disadvantages of each of these alternative approaches.

Examples are included where available to illustrate means by which states and local school districts have grappled with these issues and what has been learned from their efforts. These examples are not even, however. In some cases, we present several examples of states or districts that have implemented a particular approach, while in other cases we discuss the merits of a potential approach but have no specific examples of places where it has actually been tried. These differences reflect the fact that the field itself is uneven and is still evolving.

It is clear that states and districts are trying a wide range of experiments to identify fair and effective ways to assess the performance of all teachers, not just those who teach core subjects. Eventually, with careful documentation and evaluation of these attempts to implement performance pay, we hope to reach consensus about optimal approaches. At present, however, there is no best solution. Each potential approach entails some difficulties, and teachers and other experts simply do not agree about many of the solutions presented. Nevertheless, we are inclined to discount some objections, accept certain problems, and advocate for certain solutions—such as state-led development of end-of-course tests to measure the performance of high school teachers and English language development tests to measure the performance of ELL instructors. Until these types of measures are available, we should reward those teachers that we can based on individual measures and others via school rewards because getting incentives in, even imperfectly, is needed to change outdated teacher compensation systems that do not reward outcomes and results.

TEACHERS OF NON-TESTED SUBJECTS

Why is it difficult to measure the effectiveness of teachers of non-tested subjects?

As members of the advisory group specified, it is easy to believe that we can assess whether students read well or solve math problems well or understand social studies or science, but it is much more difficult to imagine how to assess whether students properly understand a subject such as art. Until we can agree on what constitutes effective teacher performance, it will be difficult to measure it and reward it.

Relatively little research has been conducted on teacher effects on student achievement and other educational outcomes in non-core subjects. Researchers have been largely restricted by the types of achievement test data that are available to them. Consequently, studies that have examined the effects of teacher performance-pay programs on student achievement are primarily limited to mathematics and reading/language arts (e.g., Boozer; Cooper and Cohn; Koedel; Muralidaran and Sundararaman; Ritter et al.) or to mathematics alone (e.g., Aaronson et al.; Barnett et al.; Winters et al.).⁴ Only a few studies have analyzed student test scores in other core subjects, such as science (e.g., Atkinson et al.) or science and history (e.g., Figlio and Kenny).⁵ Virtually no empirical evidence exists on teacher effects on student achievement in other subject areas.

Another complicating factor is that studies that have used student achievement test scores to measure teacher effectiveness often find substantial differences in teacher effects by subject area. Ballou proposes as an explanation that students' prior school experiences may influence their performance in some content areas more than in others. As a result,

he argues, it is more difficult in some subjects than in others to obtain reliable estimates of teachers' contributions to their students' performance:

*"For reasons probably due to the home environment, more of the variation in student reading performance is independent of school quality than is the case in math performance. As a result, it is harder to detect particularly strong (or weak) performance by reading instructors than by math teachers. . . . A recent investigation of achievement in one large Tennessee school district . . . has found that 20 percent of math teachers are recognizably better or worse than average by a conventional statistical criterion. By the same criterion, the percentage falls to 10 percent in language arts instruction and to about 5 percent among reading teachers."*⁶

Finally, it is particularly difficult to measure the productivity of teachers of non-tested subjects because we do not yet know with any degree of certainty the extent to which teachers of non-core subjects contribute to gains in student achievement in other content areas. Koedel has noted that,

*"Although the teacher-quality literature has generally assumed that same-subject teacher quality affects student performance (i.e., math teachers affect math performance and English teachers affect reading performance), it has also been implicitly assumed that off-subject teacher quality does not (i.e., math teachers do not affect reading performance)."*⁷

But there is little empirical support for this assumption, as Koedel points out:

*". . . there has not been any research to identify which teacher types (i.e., math, English, science, etc.) affect which educational outputs in secondary school. Similarly, the extent to which teachers across subjects are complements or substitutes in the production function is also unknown. . . . For example, teacher quality in some subjects may spill over into student performance in others. These spillover effects should perhaps be incorporated into teacher evaluations, but only for relevant teacher-subject matches and only if they are properly weighted. Otherwise, teachers' incentives would be poorly aligned with performance and free-riding opportunities could be enhanced."*⁸

Koedel's own study is one of the first to produce empirical evidence on the extent to which teachers contribute to gains in student achievement in other content areas. Koedel analyzed Stanford 9 test scores for high school students in the San Diego Unified School District from 1997-98 through 2001-02 to examine the effects of multiple teacher inputs on student achievement gains in reading and mathematics. He found that social studies teachers—not just math teachers—contributed to student gains in mathematics achievement. In addition, he found that math teachers—not

just English teachers—contributed to student gains in reading. This suggests that rewarding only the math teachers when math scores improve, or rewarding only the reading/English language arts teachers when reading scores improve, may unfairly penalize other teachers who also contribute to student academic growth.⁹

What are some potential ways to measure the productivity of this group of teachers and what are the advantages and disadvantages of each of these approaches?

A recent analysis of state-supported teacher pay-for-performance programs found that an element common across the nine programs examined was that all classroom teachers were eligible for some type of performance awards.¹⁰ However, examination of a broad array of state and local performance-pay programs indicates that there is wide variation among programs with regard to the type of rewards that teachers of core and non-core subjects can earn (e.g., individual vs. group), the maximum size of the rewards that the two groups of teachers can earn, and the measures that states and districts use to determine the effectiveness of core and non-core teachers. This section describes four different options that are currently in use across the United States. The options presented are as follows:

1. Both core and non-core teachers are eligible for schoolwide performance awards only.
2. Non-core teachers are eligible for some, but not all, of the individual performance incentives that teachers of core subjects can earn.
3. No student test scores are used to determine non-core teachers' eligibility for rewards. Instead, eligibility is based exclusively on non-test measures, such as observed evaluations of classroom performance, acquisition of additional knowledge and skills, or assumption of additional roles or responsibilities.
4. States or school districts adopt or create new student tests to assess teacher performance in the non-core subjects.

Option 1

Both core and non-core teachers are eligible for schoolwide performance awards only.

Examples of state-supported performance-pay programs that offer teacher bonuses based on schoolwide performance only are Alaska and North Carolina. In each case, the state sets performance improvement targets for individual schools and provides bonuses to all teachers when the school meets or exceeds those targets.

In Alaska, for example, the state's Public School Performance Incentive Program (PSSIP) rewards all teachers and support staff in schools that maintain current high levels of student performance in reading, writing, and math and in schools that make greater than expected student achievement gains on the state achievement test in these core subjects. Schools with any combination of grades except those that enroll 11th and 12th graders only are eligible to earn incentive pay.

Each student's standardized test scores in reading, writing, and mathematics are placed in one of seven performance categories. Students earn points for the progress that they make on the test from year to year and for moving to a higher performance category. The points are then used to compute an average school improvement score. A growth index developed by the state department of education assigns the school's improvement score to one of four levels.

The index was designed so that a score of 100 is considered one year of growth, the amount of improvement that all students are expected to make in one year. Schools that earn average scores of 100 or less receive no bonuses. However, all instructional and support staff in schools with average scores that exceed 100 on the growth index receive bonuses; the higher the score, the bigger the bonus. Certified staff, which include teachers, administrators, counselors, librarians, and nurses, receive bonuses ranging from \$2,500 to \$5,500. Noncertified staff such as paraprofessionals, administrative support staff, and custodians receive bonuses as well, although the size of the bonuses is smaller, ranging from \$1,000 to \$2,500:

School's level of improvement		Bonus	
		Certified staff	Noncertified staff
Strong	(107 – 108.99 point increase)	\$2,500	\$1,000
High	(109 – 111.99 point increase)	\$3,500	\$1,500
Excellent	(112 – 114.99 point increase)	\$4,500	\$2,000
Outstanding	(115 point increase or higher)	\$5,500	\$2,500

Basing the bonuses of non-core teachers on schoolwide performance in the core subjects does have a number of distinct advantages. First, the reward structure emphasizes that all teachers in the school contribute to student achievement gains in reading, writing, and mathematics. Second, it ensures that performance awards are aligned with one of the state's highest priorities, improvement in core academic subjects. Third, many teachers view this type of performance-pay program as a fair and collegial way to compensate teachers because all benefit when student performance improves. This approach may also be more in line with the cultural preferences of the community, particularly American Indian/Alaska Native communities. This approach also avoids the expense of creating new evaluation instruments or tests in non-tested subjects.

States and districts must weigh these advantages against several potential disadvantages, however. Goldhaber notes that one of the biggest disadvantages of schoolwide awards is that there may be little incentive for the least effective teachers to improve their performance if they are already assured of receiving the same rewards as the top-performing teachers in the school.¹¹

Another potential disadvantage to consider is that some research suggests that workers emphasize the parts of their jobs that are most highly rewarded.¹² If non-core teachers' bonuses depend heavily or exclusively on student gains in math or reading, the reward structure could encourage them to spend less time on their own subject and more time on the subjects in which student achievement gains are financially rewarded. In one review of the literature, Koretz cites multiple studies documenting that teachers tend to reallocate instructional resources and deemphasize materials from the same subject or a different one in response to high-stakes testing demands.¹³

TeacherSolutions particularly dislikes basing bonuses for teachers of non-core subjects on schoolwide performance in more easily measured subjects, stating that "we also question the wisdom of tying student performance in reading to teacher rewards in all non-tested content areas, as some districts have proposed."¹⁴ Although a performance-pay system of this type is convenient to administer, it does not encourage all teachers to demonstrate excellence in their area of specialization, nor does it reward teachers for their own knowledge, skills, and abilities. As one teacher argued,

*"How would you like it if YOUR bonus was based ONLY on what other teachers did or did not do? Put yourself in the shoes of someone who benefited just because of the efforts of others. That doesn't even make sense! What a meaningless reward. I want to be valued because of what I do, not what someone else does."*¹⁵

Variation on Option 1: Non-core teachers decide whether they will emphasize math or reading skills in their classes and thereby choose whether their rewards will be based on schoolwide math or reading score gains.

An unusual variation of Option 1 is found in South Carolina schools that participate in the Teacher Advancement Program (TAP), a teacher development strategy that includes performance-based compensation. As a general rule, TAP programs use three criteria to assess teacher performance: the classroom achievement gains of the teacher's own students; schoolwide achievement gains; and periodic classroom observations by trained evaluators who use TAP evaluation rubrics to assess the teacher's skills, knowledge, and responsibilities.

Rather than rewarding all teachers in a school when schoolwide performance in core academic subjects improves, Chait notes that in South Carolina, "specialists are given the option to reinforce either math or reading and be evaluated based on student gains in one of those subject areas. . . ." ¹⁶ Teachers still have the option of being evaluated based on a 50/50 split of teacher observations and schoolwide growth.

Similarly, teachers of non-tested subjects in Orange County, Florida can opt to earn performance rewards by linking themselves to teachers of tested subjects who have set student achievement goals. Both groups of teachers earn bonuses if the teachers of tested subjects meet their targets. According to Milanowski, "The district believes this will encourage teachers in non-tested subjects to support the efforts of their colleagues, as well as provide a way for these teachers to earn the bonus."¹⁷

An advantage of these variations is that they give teachers more choice and control over the alignment between their pay and their performance. Yet as previously noted, very limited research has been conducted on the extent to which non-core teachers contribute to gains in student achievement in other content areas. Koedel found that only some groups of teachers contributed to student gains in mathematics (i.e., math and social studies, but not science or reading/English language arts), and only some groups of teachers contributed to student gains in reading (i.e., reading/English language arts and math, but not science or social studies).¹⁸ Although allowing teachers to tie part of their pay to student gains in another academic subject of their choice is an interesting idea, states and districts are advised to proceed with caution, given the limited evidence that is currently available on the effectiveness of this practice and the relative merits of an approach that is decidedly more complicated to administer than rewarding all teachers when schoolwide math or reading performance improves.

**Option
2**

Non-core teachers are eligible for some, but not all, of the individual performance incentives that teachers of core subjects can earn.

An example of a district that has tried several different versions of this option is Houston. In 2000, the Houston Independent School District adopted a performance-pay system based on schoolwide improvement in reading and mathematics on the state achievement test. In 2006, the district began awarding individual as well as schoolwide bonuses, in part because the previous compensation system failed to reward highly effective teachers who taught at low-performing schools.

The new performance-pay system that the district adopted enabled teachers of core academic subjects in all schools, not just high-performing ones, to earn bonuses based on the learning gains made by students in their own classrooms. At the same time, all teachers and staff qualified for bonuses if schoolwide performance improved. During the first year of implementation, core teachers could earn up to \$6,000 in bonuses, and non-core teachers could earn up to \$3,000 in bonuses under three strands:

- 1. Strand 1:** Bonuses were based on schoolwide improvement in reading and math on the state's criterion-referenced achievement test, the Texas Assessment of Knowledge and Skills (TAKS), when compared to similar schools across the state. All faculty and staff in the school were eligible for bonuses.
- 2. Strand 2:** Bonuses were based on individual student progress in the core subjects assessed on the norm-referenced Stanford (English) and Aprenda (Spanish) achievement tests. Core teachers received performance awards under this strand if their students made greater learning gains than students in other schools across the district with similar demographics. Teachers of non-core subjects were also eligible for bonuses, but the maximum amount that they could earn under this strand was half the amount that core teachers could potentially earn.
- 3. Strand 3:** Bonuses were based on year-to-year student progress in reading and math on the state TAKS test. Only core teachers were eligible to earn this type of bonus.¹⁹

Many features of the new reward system were controversial, but differences in the amounts that core teachers could earn, compared to those who taught non-tested subjects or grades, generated some of the harshest criticism from teachers when Houston made its first payout in 2007. Comments made by some Houston teachers indicate that the new reward structure presented a major shift in culture for many educators in the district:

“My problem with this system is that non-core teachers received more money than some classroom teachers. I agree that they help us with our kids but they only spend 45 minutes a week with them. We have them at least 30 hours a week. Why did we not get at least the same amount as them?”²⁰

“...I am an ART teacher and did not get one cent. I only qualified for Strand 1, and because I don’t teach standardized tests I don’t even get considered for the other money, so I don’t really know how many other “ancillary staff” made much money at all with this.”²¹

“...the distribution of the large bonuses to only those teachers who teach a TAKS related class is an injustice. Example: an 8th grade History teacher could receive a large bonus for his or her work, but the 7th and 6th grade history teachers would never receive comparable sums. This fails to acknowledge these teachers’ efforts. Teaching is a community affair—no one teacher does it alone.”²²

One of the ways in which Houston responded to this type of feedback was to hire a communications firm to help develop a district-wide communication plan so that all teachers and staff understood which teachers qualified for rewards under the different strands. Houston also made some changes to the reward structure itself, such as redesigning Strand 3 so that core and non-core teachers alike earn the same awards for improvements in schoolwide performance. Strand 2 teacher awards, which are based on individual student progress only in core academic subjects, are now restricted to teachers of those subjects. According to district staff, separating the groups and clarifying eligibility requirements has reduced criticism and concerns about the reward structure that surfaced after the first payout.²³

Dallas is another example of a district that restricts some bonuses to teachers of core subjects only. All professional staff and support staff in Dallas are eligible to earn school-based awards based on the amount of gain that students make compared to students in schools with similar demographics.²⁴ These school-based awards range from \$1,250 to \$2,000 for professional staff and from \$625 to \$1,000 for support staff. However, only teachers of tested subjects (i.e., English/language arts, mathematics, social studies, science, world languages, and computer science) are eligible to earn classroom-level awards. If student gains in their classrooms exceed the amount of gain that similar students make in other classes, these teachers can earn additional bonuses ranging from \$2,000 to \$8,000.

Like Houston and Dallas, the Guilford County, North Carolina Public Schools created a performance-pay system that targets some bonuses and performance awards to teachers of core subjects only, though the two Texas districts and Guilford County have very different goals.²⁵ Houston and

Dallas created performance-pay programs to reward teachers for improved student performance, while Guilford County originally designed its Mission Possible program to recruit and retain highly qualified teachers and effective school leaders in the district's most difficult-to-staff schools. Within those schools, the district targeted its financial incentive program to teachers of hard-to-fill subjects, especially middle and high school mathematics teachers. The district offers recruitment/retention incentives to teachers in these schools, as well as performance incentives based on student gains on the state achievement test.

Guilford County's recruitment/retention incentives are structured so that teachers in highest demand, such as algebra teachers, are eligible for the biggest incentives:

Grade level	Subject	Incentive
K-5	Elementary education	\$2,500
6-8	Language arts, reading, or math	\$2,500
6-8	Math	\$9,000
9-12	Math	\$9,000
(any grade)	Algebra I	\$10,000
(secondary)	English I	\$2,500

In order to receive a performance award, teachers must teach a tested course in a Mission Possible school. The courses must be tested either at the end of the grade (e.g., reading or math in Grades 3 to 8), or at the end of the course (e.g., algebra). Core subject teachers can receive a performance incentive of up to \$2,500 for a class that has a value-added score between 1 and 1.5 years of growth, and up to \$4,000 if the class has a value-added score that exceeds 1.5 years of growth. Teachers of non-tested subjects, such as fine arts and foreign languages, cannot receive either the district recruitment/retention or performance incentives. However, they can still receive schoolwide performance awards through North Carolina's state-sponsored incentive program if the entire school meets its academic performance targets in reading and math.

According to district staff, the recruitment/retention incentives seem to be attracting a larger pool of highly qualified applicants to high-need Mission Possible schools. Staff note that the main problems that they have encountered have been low teacher morale in some of the schools and some teacher complaints about the substantially higher incentive offered to algebra teachers. Like Houston, Guilford County has responded by improving communications with teachers.²⁶

As these examples show, creating a pay system in which some groups of teachers are eligible for rewards that others are not can be risky, and policymakers who choose this option should anticipate that teachers may have strong objections to it. Lessons learned in Houston and Guilford County

suggest that these risks can be minimized by establishing clear and consistent communications with teachers and by designing a reward structure that places core and non-core in separate eligibility categories. It is important to note that both Houston and Guilford County have educators associations, but neither has a teachers union. It is not clear that performance-pay systems that differentiate teacher pay by core and non-core subjects, would be accepted in districts that require collective bargaining.

**Option
3**

Student test scores are not used to determine non-core teachers' eligibility for rewards. Instead, eligibility is based exclusively on non-test measures, such as observed evaluations of classroom performance, acquisition of additional knowledge and skills, or assumption of additional roles or responsibilities. All of these, and other non-standardized test measures, may be displayed in a thoughtful portfolio.

An example of a school district that tried this approach is the Colonial School District in suburban Philadelphia. During the 1999-2000 school year, Colonial implemented a mandatory performance-based pay system for all of its classroom teachers, as well as some groups of non-teaching staff. The district budgeted \$200,000 for bonuses, half of which would be used to reward groups or teams of teachers. The other half would be used to reward the top-ranked 10 to 20 percent of individual teachers whose students scored well on standardized achievement tests. Teachers could receive up to \$2,500 for an individual bonus, as well as a group or department bonus of up to \$2,500.²⁷

Individual rewards for reading and math teachers were based only on their students' scores on either the statewide assessment, administered in grades 5, 8, and 11, or the Terra Nova, which was administered in grades 3, 4, 6, and 7. No test scores were used to determine teacher eligibility for bonuses in other subject areas. Instead, the district hired a consultant to identify appropriate criteria and alternative sources of input to judge individual teacher performance. The district also developed a separate evaluation system to assess the performance of teacher groups by grade level, team, and department at the elementary, middle school, and high school levels.²⁸ In the case of non-teaching staff, such as librarians, guidance counselors, and school nurses, the district based performance rewards on student and parent surveys.²⁹

A thoughtful portfolio approach is another potential way to measure teacher performance in non-tested subjects. Portfolios could contain any number of artifacts that would be useful for assessing performance in a teacher's area of specialization, such as lesson plans, videotaped lessons, and examples

of student work. States and districts might also consider incorporating some elements of the National Board for Professional Teaching Standards' national certification process. The National Board has developed detailed rubrics for assessing teacher performance in 16 subjects and at different grade levels. In South Dakota, for example, teachers select one of the nine elements of National Board Certification and receive a bonus based on successful demonstration of competencies included in that element.

A potential advantage of using non-test measures to judge individual teacher performance in non-tested subjects is that it can encourage all teachers to demonstrate excellence in their area of specialization. Another advantage is that it avoids the time and expense required to develop, administer, and score new student assessment instruments in subjects that are not currently tested. However, developing new evaluation rubrics or other non-test measures and training principals and other evaluators to use them may be just as expensive and labor-intensive, possibly even more so, than designing new tests.

Policymakers should also bear in mind that non-test measures are also imperfect measures of teacher performance. For example, research suggests that principals are quite accurate at identifying very high and very low performers, but not as accurate at assessing teacher performance in the middle of the range.³⁰ Research also suggests that there is a tendency for principals to be overly lenient with their ratings and to give higher than deserved performance evaluations, which will fail to differentiate pay over time.³¹ These results reinforce the need to use multiple measures of teacher effectiveness when making high-stakes compensation decisions.

A potential advantage of using non-test measures to judge individual teacher performance in non-tested subjects is that it can encourage all teachers to demonstrate excellence in their area of specialization.

Variation on Option 3: Non-core teachers design their own performance goals in their own area of specialization, and their rewards are based on attainment of these goals.

This variation on Option 3 can be found in Denver, as well as in Orange County, Florida. One of the multiple ways in which teachers can earn performance-pay under Denver's ProComp system is by meeting their own individual performance goals. Teachers work in collaboration with administrators to set these performance goals, which are linked to student learning outcomes. This approach may be particularly appealing to teachers of non-tested subjects, because they can design performance targets specific to their subject area and earn rewards for achieving them. TeacherSolutions supports this option as a reasonable way to assess the performance of other educators in a school whose contributions to student achievement are difficult to measure with standardized achievement tests, such as guidance counselors, librarians, and teacher specialists who work with students in other teachers' classrooms.³²

Similarly, middle and high school teachers in ten Orange County, Florida schools can choose to set student achievement goals, but only for subjects tested on the statewide Florida Comprehensive Achievement Test (FCAT).³³ Participating teachers work with their principals to set goals for improvement in student proficiency levels in reading, writing, mathematics, and/or science, based on their students' test scores the previous year. Successful participants can earn bonuses ranging from \$1,000 to \$4,000, depending on the percentages of their students who make the required amount of gain. Participants must also receive satisfactory evaluations and complete required subject-specific professional development.

One of the strengths of allowing teachers to earn rewards by setting and meeting individual performance goals is that it gives teachers a choice in how their performance is to be assessed, and it is clear to teachers what they must do to earn an award. This approach also encourages all teachers to excel in their own area of specialization. It avoids the free-rider problem because only the individual teacher is responsible for meeting the goals, and it also avoids the need to develop new assessments in non-tested subjects. However, this approach will require thoughtful planning and clear direction to teachers and administrators so that performance goals are set at appropriate levels of difficulty and are measurable.

**Option
4****Create new student tests to assess teacher performance in non-core subjects.**

Florida is the best-known example of a state that has opted to develop new assessments as a way to measure teacher effectiveness in non-core subjects. Florida has used this approach since 2006, when it adopted Effectiveness Compensation, or E-Comp, the first of three consecutive state performance-pay plans.³⁴

E-Comp required districts to pay bonuses of at least 5 percent to teachers who had satisfactory evaluations and ranked among the top 10 percent of teachers in the state, based on annual student achievement gains. Reading and math teachers were to be judged exclusively on student improvement on the FCAT (Florida Comprehensive Assessment Test). For teachers of other subjects, districts were required to develop or adopt standardized measures of performance based on student learning gains, though not necessarily paper-and-pencil tests.³⁵ The new assessments would be used to track student progress and determine teacher bonuses, but would not affect high school graduation or the annual letter grade ratings that Florida assigns to each school, based on state accountability measures.³⁶

This directive from the state presented a host of challenges for school districts. When the state adopted E-Comp in February 2006, less than one-third of Florida's teachers (31 percent) taught reading and math. Districts were given a deadline of June 15th to submit proposals describing how they planned to evaluate the remainder of their teachers who taught non-FCAT-tested subjects. New compensation plans were to be ready and in place in fall at the beginning of the 2006-07 school year.³⁷ District superintendents argued that this was an unrealistically short amount of time to prepare a thoughtful and feasible plan, let alone develop new assessments and overhaul existing compensation systems.³⁸

Until new assessments in other content areas could be developed, the state department of education directed districts to use annual performance evaluations to identify top-performing teachers of non-tested subjects. A problem with this interim solution was that districts used different evaluation scales, which made it difficult to judge the top-performing 10 percent of teachers across the state in a fair way.³⁹ Moreover, in districts such as Miami-Dade, teachers were simply rated as "satisfactory" or "unsatisfactory," and any changes to the evaluation system would be subject to lengthy negotiations with the teachers union.⁴⁰

In April 2006, the Florida legislature replaced E-Comp with the Special Teachers Are Rewarded, or STAR, compensation system. STAR gave districts additional time to develop their teacher evaluation plans and increased the

proportion of teachers eligible for rewards from the top 10 percent to the top 25 percent. Instead of relying exclusively on achievement test scores to evaluate teachers, districts could base up to 50 percent of a teacher's evaluation on principal observations. In addition, STAR allowed districts to use standardized tests other than the FCAT to assess the performance of reading and math teachers. Because few districts had secondary-level tests in science and social studies, districts were also allowed to base performance incentives for secondary science and social studies teachers on FCAT reading scores, rather than content-area assessments.

For teachers of non-core subjects, districts were to use a district-wide assessment based on state standards in the subject that the teacher taught. Very few districts had tests that met these criteria, however. Fortunately, Hillsborough County had already developed several hundred end-of-course examinations and agreed to host a clearinghouse established by the state department of education to share these tests with other districts and schools across the state.⁴¹ The Florida Virtual School also agreed to contribute exams to the clearinghouse. According to Hillsborough County officials, school districts and individual public schools used the clearinghouse website more than 16,000 times by February 2007 to download the exams.⁴² But some school systems were reluctant to use other districts' tests to evaluate their own teachers. In Pasco County, for example, "school Board members particularly balked at the idea of borrowing tests from other districts so they could determine how Pasco teachers' classes were doing."⁴³

Even when districts were willing to use existing tests, they often had no pretests to measure student learning gains over time, as prescribed by the state. The state had proposed that districts award points to teachers when student performance reached higher proficiency levels on the FCAT, and the state developed a formula to determine how to award points for student growth in reading and math at more difficult and less difficult levels. But without pretests in other subjects to measure student growth, districts could not assess the performance of the vast majority of their teachers who taught subjects other than reading and math. By November 2006, the state was still developing non-FCAT assessment instruments.⁴⁴ A highly unusual—and difficult to defend—solution was proposed for the first year only (2006-07), which was to allow districts to use the FCAT reading test as the pretest measure for non-core subjects, such as a Grade 11 State Government course.⁴⁵

In March 2007, the state legislature replaced the STAR compensation system with the current Merit Award Program, or MAP. MAP placed even greater emphasis on student achievement as the primary criterion for earning performance pay, but in many ways gave districts more local control. Unlike STAR, district participation in MAP was voluntary. MAP also

allowed districts to base teacher performance awards on either student learning gains or proficiency levels and to reward teacher teams, not just individuals. Moreover, districts could increase the size of the awards and decide the proportion of teachers to be rewarded.⁴⁶ But districts could no longer use the FCAT reading test to measure the performance of teachers of non-FCAT subjects and grades. Instead, participating districts had to develop or adopt district-wide assessments in each content area.

It is not surprising that a number of serious testing errors were discovered during the transition from STAR to MAP, given the pressure that districts were under to develop and administer new instruments quickly. Orange County, in particular, was criticized in the local media for producing hastily constructed tests that contained hundreds of errors, including a German test partially written in French.⁴⁷ Most of the flawed tests were given in high school elective courses, and teachers of these courses were disproportionately less likely than other teachers in the district to win performance awards. Though they comprised 35 percent of the teachers who were eligible to win awards, only 7 percent earned bonuses.

Several other Florida school districts chose to opt out of MAP after discovering that it would cost several hundreds of thousands of dollars in local funds to develop the content-area exams that they would need to comply with state requirements. Hernando County officials, for example, estimated that designing new tests and teacher evaluation systems for MAP would cost the district almost \$400,000.⁴⁸ Seminole County officials estimated that they would need to spend at least \$300,000 to develop more than 500 tests, and Volusia County estimated that its costs could be \$600,000 or higher.⁴⁹

One of the key lessons learned in Florida is that to maximize acceptance, designers of alternative compensation systems need to pay close attention to teacher perceptions of fairness. In Orange County, flawed and hastily constructed tests received much of the blame for the fact that teachers of high school electives were disproportionately less likely than other teachers to earn performance pay. Moreover, research cited previously suggests that it is easier to identify top-performing teachers in some subjects (e.g., math) than in others (e.g., reading and language arts). As members of the advisory group noted, this would suggest that mixing math and reading teachers together would likely lead to a disproportionate share of math teachers identified as the top performers. Under E-Comp, Florida districts did rank the top 10 percent of math teachers separately from the top 10 percent of reading teachers, but they were directed to pool the rest of their teachers together to determine the top 10 percent of non-FCAT teachers across the state. One of the changes made under the subsequent STAR plan was to allow districts either to compare teacher performance across subjects or to select the top 25 percent of teachers within each subject.⁵⁰

One of the key lessons learned in Florida is that to maximize acceptance, designers of alternative compensation systems need to pay close attention to teacher perceptions of fairness.

The most important lesson learned from Florida is that creating new tests to assess teacher performance in all non-core subjects is a very complicated, time-consuming, and expensive task. As these examples from Florida show, unrealistic timelines are likely to lead to indefensible strategies, such as tying teacher pay to “growth” between a reading pretest and a posttest in an entirely different subject area. Examples from Florida also show that unanticipated costs of developing and administering new tests can, and do, prompt districts to opt out of state performance-pay programs. Good quality testing is always costly, whether in development of end-of-course tests or large-scale assessments, but one way to hold down costs would be to limit the number of new tests to the most critical subjects. Another way would be to establish a state clearinghouse, as Florida did, so that districts and schools can share assessments and avoid duplication of effort.

Taken together, the options presented above provide viable ways for teachers of non-tested subjects to be more fully integrated into a performance measurement and award system. It is important to note that states and districts interested in designing a performance-based pay system that includes all teachers must consider not only how they will evaluate the performance of teachers of non-tested subjects, but how they will evaluate the performance of teachers of non-tested grades. This issue is addressed in the following section.

TEACHERS OF NON-TESTED GRADES

Why is it difficult to measure the effectiveness of teachers of non-tested grades?

Teachers of pre-kindergarten to Grade 2 and high school teachers are two groups that pose specific measurement challenges when designing a performance-based compensation plan that includes all teachers. One of these challenges is a lack of test score data at both ends of the continuum. In the early grades, lack of test score data precludes the use of value-added measurement as a way to assess teacher productivity because value-added measurement experts recommend analyzing at least 3 years of standardized achievement test data to determine expected gain. However, most districts do not administer standardized achievement tests before Grade 3.

Administering standardized achievement tests in the earlier grades could be done, but early childhood education experts strongly caution policymakers not to use these types of paper-and-pencil-type tests to assess what young children know and can do.⁵¹ Existing achievement tests, they argue, are not developmentally appropriate for use with children below Grade 3. Though technically feasible, it would be unwise to use such tests as a measure of teacher effectiveness in pre-kindergarten to Grade 2 for purposes of teacher compensation.

Although assessment instruments that are appropriate for the early grades do exist, they tend to be adaptive tests designed to provide diagnostic feedback for instructional purposes, such as identifying children for early intervention or special education services, not to assess academic content mastery. Examples of assessments that are commonly used for this purpose are the DIBELS (Dynamic Indicators of Basic Early Literacy Skills) and the MAP (Measures of Academic Progress). Whether it is appropriate to use these tests to measure teacher productivity is not yet known because there is no solid research on the technical adequacy of using diagnostic tests for this purpose. Moreover, the potential for measurement error is high because these assessments rely on very few observations.

High school presents a different set of measurement issues. One of the main challenges is that many of the tests that are administered in high school, such as high school exit exams or minimum competency exams, are not particularly helpful measures of how effective teachers were at increasing student learning in their particular classes:

“The main problem is that student outcomes cannot be attributed to the performance of particular teachers because the material covered on such tests goes well beyond that covered in a specific course. In addition, such tests can shed little light on how effectively teachers succeed in conveying high school-level material because the material covered is often at a relatively low level—one more appropriate to the middle school than to the high school. What is needed, instead, are tests that are external to the school, that relate to the material that teachers are hired to teach, and that the students are likely to take seriously.”⁵²

Another measurement challenge is that there is no defined scope and sequence of curriculum at the high school level as there is at the elementary level. For example, there is no secondary “science” test that adequately covers all topics in science. In one school, earth science may be required in 9th grade, followed by biology in 10th grade, chemistry in 11th grade, and physics in 12th grade. In other schools, students may opt out of these courses, take them in different sequence, or take courses in core subjects in alternating years. One researcher noted that in the district that he studied, “the typical student . . . alternates between taking science and social studies in the 9th and 10th grades and, generally speaking, only takes these subjects concurrently in the 11th grade.”⁵³ When such radical differences exist in the delivered curriculum, it is difficult to interpret standardized test results to measure and reward teacher performance. As Koretz points out, the peculiarities of the high school curriculum make it particularly difficult to use a value-added approach, because value-added measurement systems require

“at least annual testing using assessments with content that overlaps substantially from grade to grade. If testing is done only annually, this approach is useful only when the curriculum is cumulative across grades. For example, it could be useful in reading and probably in elementary school mathematics, but its applicability to secondary school science is arguable.”⁵⁴

An additional measurement challenge is that selection effects are more pronounced at the high school level because students are not randomly assigned to teachers. The most effective teachers may opt to teach the most advanced classes, or principals may decide to assign the most effective teachers to low-performing students who are most in need of an effective teacher. In addition, some high school teachers may teach only in grades that do not administer statewide achievement tests, such as 11th- and 12th-grade calculus. An individual teacher reward system would exclude these teachers, who may very well be some of the top teachers in the school.

Finally, administration and scoring of any instrument intended to evaluate teacher effectiveness, whether it is a test, rubric, or principal evaluation, is subject to corruptibility. It is important to keep in mind that the incentive for a teacher to corrupt the administration or scoring of teacher evaluation instruments is quite high if the results are tied to pay. This is not to imply that teachers will necessarily cheat or otherwise attempt to alter evaluation results in their favor. Rather, the point is that the person who is being evaluated should not be charged with the responsibility of administering or scoring the instrument, whether it is the DIBELS, an end-of-course test, or any other measure, for purposes of determining compensation.

What are some potential ways to measure the productivity of teachers in pre-kindergarten to Grade 2, and what are the advantages and disadvantages of each of these approaches?

States and districts that wish to include teachers in the lower elementary grades in their compensation plan tend to choose among three options:

Option 1

Create a developmentally appropriate rubric to assess how well teachers in pre-K to Grade 2 are supporting young children’s development on dimensions such as their cognitive development, social-emotional development, motor development, and language acquisition.

The Colonial School District in suburban Philadelphia used this approach in 1999-2000. Beginning in Grade 3, teacher performance was assessed on the basis of students’ standardized test scores, but “pupils in lower grades were measured based on more subjective criteria, including “competence in the social/emotional domain,” “mastery in academic areas,” and “competence in the psychomotor domain.”⁵⁵

As another example, researchers at the Center for Advanced Study of Teaching and Learning at the University of Virginia developed the Classroom Assessment Scoring System (CLASS) as a tool to evaluate teachers of students in pre-kindergarten and the lower elementary grades.⁵⁶ It includes observations of classrooms by trained evaluators who measure teachers' performance in three areas: emotional support (including classroom climate, teacher sensitivity, and student perspectives), classroom organization, and instructional support (including quality of feedback and language modeling). Massachusetts and Wyoming currently use CLASS to assess and provide support for preschool teachers in and as part of a web-based professional development system called MyTeachingPartner. The American Board for Certification of Teaching Excellence is using the system in a pilot program to recognize distinguished teachers.

**Option
2**

Use student results from adaptive tests such as the DIBELS or MAP to assess teacher performance at the early grades.

Using student results from the DIBELS, the MAP test, or a similar assessment is one potential way to measure the performance of teachers of pre-K to Grade 2, but it is risky because of current limitations of research on the technical adequacy of using such tests to measure teacher performance. One caution that members of the advisory group had about the use of adaptive tests pertained specifically to students with disabilities. advisory group members pointed out that students with disabilities often have splintered skills, so that they may be quite competent at some higher level skills that are tested on adaptive tests but not the lower level skills. Because of the adaptive nature of the test, the students never get to show the full range of their competence. If they cannot answer some of the basic questions, they are directed to easier items. In such a case, the test may underestimate teacher productivity and decrease the teacher's chances of earning a performance award.

**Option
3**

Use measures other than individual classroom achievement as a way to include teachers of pre-K to Grade 2 in the compensation system.

An example of this strategy can be found in Guilford County, North Carolina. Teachers of kindergarten to Grade 2 classes in the district's Mission Possible schools are not eligible to earn the same performance incentives as other teachers because the state prohibits testing below Grade 3. After the district made its first payout in fall 2007, school board members pressed district staff to find a way to reward early elementary teachers as well.⁵⁷

Guilford County's solution is to include K-2 teachers in its recruitment/retention incentive program, which allows them to earn \$2,500 each year

for teaching Grades K-2 in a Mission Possible school.⁵⁸ The district also offers a non-financial incentive to these early elementary teachers—a smaller than average class size of 15 students per teacher.⁵⁹ In addition, the teachers can qualify for schoolwide bonuses through North Carolina's state-sponsored incentive program. If the school meets its student achievement performance targets, every teacher in the school, including those at the early elementary grades, earns a bonus.

What are some potential ways to measure the productivity of high school teachers, and what are the advantages and disadvantages of each of these approaches?

States and districts that include high school teachers in performance-based compensation systems tend to choose one or more of the following strategies: use existing tests that were intended for other purposes; buy or develop new tests; move to a group reward structure at the high school level, rather than offering individual teacher rewards; or supplement student test outcomes with schoolwide measures such as high school dropout and graduation rates.

Option 1

Fill in non-tested grades by using existing tests that were developed for other purposes.

One way to fill in data for the high school grades that do not administer a statewide achievement test is to purchase off-grade tests from the same vendor that the district uses to test Grades 3-8. This option solves problems of scalability. That is, it allows valid measurement of annual student learning gains because a different version of the same test is given in successive years. Another option is to use the MAP test to fill in non-tested grades, but we do not yet know whether this is an appropriate use of this test instrument, particularly in Grades 11 and 12. The test developers themselves report that “in general, students in grades 2 through 10 take MAP tests,” but they are cautious about recommending the use of the MAP above or below this grade range.⁶⁰

Student performance on college entrance examinations such as the ACT and the SAT is another source of information that states and districts often consider using as a measure of high school teacher effectiveness. But both the ACT and the SAT were designed to measure student preparedness for college, and the test makers themselves acknowledge that the tests are not appropriate measures of teacher productivity. The tests are not curriculum-sensitive because they are not aligned to state standards, and the content covered on the tests extends well beyond what is taught in a single high school course. This makes it difficult, if not impossible, to attribute student outcomes to individual teachers.

**Option
2****Adopt or develop new end-of-course tests.**

Some states, such as North Carolina, have developed statewide end-of-course examinations in key subjects such as algebra that students must pass to receive credit toward high school graduation. As previously discussed, Florida has also instituted policies that require districts to adopt or develop district-wide tests in subjects that are not already tested on the state assessment.

States and districts that choose to adopt a similar approach should consider limiting the number of tests that are developed and to allow sufficient time for test development. A number of Florida districts chose to opt out of the state performance-pay program when the cost of developing hundreds of new examinations in a very short amount of time came to light. Of course, one could argue that while developing local assessments is more costly than purchasing off-the-shelf tests, local assessments are also a better fit for local standards. Local tests can also deliver substantial curricular alignment and articulation benefits, which help to offset the development costs. A state that does decide to develop new tests should consider establishing a clearinghouse, as Florida did, to enable districts to share their work with each other, reduce costs, and prevent duplication of effort.

Whether states and districts decide to purchase off-grade tests from the same vendor, administer the MAP, or develop new end-of-course tests to assess student learning, they should carefully consider whether the results will count toward high school promotion or graduation or some other outcome that also matters to students if they plan to tie the test results to teacher compensation. States and districts should also carefully consider the degree to which the tests are aligned to the school's curricular scope and sequence and to state standards in order to judge whether the tests are accurate measures of teacher productivity.

States and districts should also carefully consider the degree to which the tests are aligned to the school's curricular scope and sequence and to state standards.

**Option
3****Base rewards of high school teachers on department-wide performance, rather than on individual classroom performance.**

Instead of adopting or creating new tests to assess individual high school teacher performance, states and districts could opt to change the reward structure instead. As previously explained, the lack of a defined curricular scope and sequence at the high school level, infrequent testing, and selection effects make it difficult to determine which teachers to credit when student test performance improves. One way to solve this problem is to base rewards of high school teachers on department-wide performance, rather than on individual classroom performance. A reward structure that

bases a portion of teacher awards on group or schoolwide gain is more likely to encourage all teachers in a school to assume responsibility for all students, as opposed to a reward structure that bases teacher awards only on the amount of gain made by their own students.

Houston, for example, decided to move to department-level awards after the district administered its first round of individual teacher payouts in January 2007. Teachers in Grades 9 through 12 are all rewarded in the same manner now. They can earn up to \$5,000 in bonuses for department progress, based on the department-level score in the subject taught. According to district staff, two of the advantages of this approach are that it encourages teachers to work together as a department to improve student achievement, and it allows the district to reward 12th-grade teachers even though 12th-grade students do not participate in state testing.⁶¹

Option 4

Supplement student test outcomes with schoolwide measures such as high school dropout and graduation rates.

Basing a portion of teacher pay on schoolwide dropout and graduation rates aligns teacher performance with two of the highest priorities of secondary school. States and districts that choose to incorporate these kinds of outcome measures in their compensation system should consider a model that rewards teachers for closing gaps in dropout and graduation rates for poor and minority students, for English language learners, and for students with disabilities. Attainment models that simply reward teachers in schools that have the lowest dropout rates and/or the highest graduation rates would likely favor schools in more affluent communities and could encourage teachers to move away from high-poverty, low-performing schools.

Post-high school outcomes could conceivably be other useful schoolwide measures of high school teacher effectiveness, such as the percentage of high school graduates who enrolled in college but had to take remedial coursework. The advisory group cautioned, however, that post-high school outcomes present formidable data collection challenges. Moreover, they cautioned that compensation systems should not be designed to pay teachers for some measure of their high school that was determined by its performance 6 or 7 years ago, particularly in schools with high rates of teacher and administrator turnover.

Although each of the approaches presented has distinct advantages, the preferred way to assess what an individual high school teacher contributes to student learning is to measure student performance at the beginning and at the end of a course. It is important that states, rather than individual districts, do the bulk of the work required to develop end-of-course assessments if incentive pay is going to work for high schools. Some nationwide tests from

vendors are available that might also be considered, but only the state can develop a comprehensive suite of tests that will apply to most subjects because it is time-consuming and expensive to do this well.

Restricting the development of end-of-course assessments to a select number of key subjects seems to be a reasonable and cost-efficient way to go about doing this (e.g., algebra, geometry, chemistry, U.S. history, etc.). Teachers of these key subjects could be rewarded for improved performance at the individual classroom level, while teachers of other high school subjects could be rewarded for improved performance based on schoolwide or department-wide measures.

The next sections examine the challenges of designing and implementing a performance-based compensation system that includes teachers of two rapidly growing student subgroups, English language learners (ELLs) and students with disabilities. Teachers in these categories include not only teachers of bilingual education, English as a second language, and special education, but also mainstream classroom teachers who have ELLs or students with disabilities enrolled in their classes.

TEACHERS OF ENGLISH LANGUAGE LEARNERS

Why is it difficult to measure the effectiveness of teachers of English language learners?

The extreme diversity among ELLs makes measurement difficult because language is only one of the educationally relevant characteristics of this student population. ELLs differ with respect to home language, language of instruction, native language proficiency, English language proficiency, prior schooling, and academic content knowledge in English and in the native language. Students with different prior knowledge and different levels of English language proficiency are likely to learn academic content in English at very different rates, even when taught by the same instructor.

ELLs, particularly new immigrants, also tend to have high rates of mobility. There is a strong probability that they will be missing achievement test scores and other demographic data that schools need to calculate year-to-year learning gains. There is also a strong probability that ELLs will attend more than one school within the same year. Comparing annual schoolwide averages in schools with high rates of student turnover is problematic because “the school’s growth targets are based on student performance from the year before, even though many of those students are no longer at the school and different students with different achievement levels have taken their place.”⁶²

A value-added approach is a preferred way to measure student growth, particularly for teacher compensation purposes, because it can distinguish among multiple teachers and schools and pro-rate achievement gains

based on dosage. However, expected gain is based on the assumption that students will perform similarly from year to year. We do not know whether this assumption applies to students who are acquiring content knowledge at the same time that they are acquiring proficiency in a second language. Even 3 years of value-added data may not be sufficient to predict 1 year of expected growth accurately for an ELL who is tested in English because we do not know how stable academic achievement gains are from year to year for ELLs.

Some states require ELLs to take academic achievement tests in English after only a brief period of English academic instruction. Research suggests that ELLs typically acquire conversational fluency in English in 3 to 5 years, but it takes an average of 4 to 7 years to develop the level of academic language proficiency required to perform schoolwork successfully.⁶³ Achievement tests administered in English will likely underestimate what students early in the process of learning English know about reading, math, science, and other subjects and will likely yield inaccurate information about teacher performance. As emphasized by the National Academy of Sciences:

“when students are not proficient in the language of the assessment (English), their scores on a test in English will not accurately reflect their knowledge of the subject being assessed” (NRC, 1999, p. 214). Therefore such assessments provide neither accurate data for accountability purposes, nor do they help teachers to enhance their instruction. As the National Research Council noted, “if a student is not proficient in the language of the test, her performance is likely to be affected by construct-irrelevant variance—that is, her test score is likely to underestimate her knowledge of the subject being tested” (NRC, 1999, p. 225 ⁶⁴).”

A more accurate assessment of teacher performance is usually obtained by supplementing student achievement measures with observed evaluations of teaching performance. Many districts have also adopted alternative assessment systems such as long-term projects and portfolio-based assessments as methods of documenting and measuring the progress of ELLs.⁶⁵ The challenge with this approach is standardization of the assessment tool and measurement of results. Any attempt to attach financial incentives to these measures should take these challenges into account.

Which teachers to reward becomes a very important question if it takes several years for students who are in the process of acquiring English to show marked improvement on English achievement tests. Though ELLs may have the content knowledge and ability to perform assessment tasks successfully, lack of proficiency in English limits their ability to demonstrate the full extent of what they know and are able to do.⁶⁶ At some point,

students may appear to make huge gains in subject matter knowledge in one year's time, when in fact they may have simply reached a level of proficiency in English that enables them to demonstrate their cumulative knowledge on an English achievement test. An individual teacher reward system might credit only the current year's teacher-of-record for learning gains, whereas a team-based approach would reward all of the teachers who had worked with the student. The prevalence of non-traditional instructional arrangements for ELLs also argues for team-based rewards (e.g., pull-out instruction, assistance from an instructional aide, native-language teacher paired with an English-speaking teacher, English as a Second Language (ESL) teacher teamed with teachers of core academic subjects).

In some districts it may be difficult to develop a uniform district-wide teacher compensation system due to variability in the concentration of ELLs across schools. In such cases, districts may have to develop a fair set of subsystems. For example, a district could assess the level of teaming in each school and then allocate some schools to team rewards and others to non-team rewards. At the same time, it would need to ensure that the target bonus amounts were similar.

A different type of measurement challenge arises in states that allow districts to provide initial academic instruction and tests in students' native languages. Over time, students transition to increasing amounts of English instruction and all-English achievement tests as their proficiency improves. But it is difficult to use value-added measures to calculate expected gain when the language of the test (and the test form itself) changes as students become increasingly proficient in English. That is, it is hard to say if we are getting a true measure of improvement by calculating gain from a 5th-grade math test given in Spanish to the 6th-grade math test given in English.

What are some potential ways to measure the productivity of this group of teachers and what are the advantages and disadvantages of each of these approaches?

It may be tempting simply to exclude ELLs from accountability systems and their teachers from performance-based compensation systems, given the formidable measurement and implementation challenges that have been described. However, it is very important to include ELLs in the assessments and include their teachers in reward systems so that all are held accountable for increasing students' language proficiency and subject matter knowledge. The performance of all students, including ELLs and students with disabilities, should be assessed and reflected in how teachers are motivated and rewarded. Three of the ways in which this can be done for teachers of ELLs are as follows.

... it is very important to include ELLs in the assessments and include their teachers in reward systems so that all are held accountable for increasing students' language proficiency and subject matter knowledge.

**Option
1****Base performance rewards for teachers of ELLs on schoolwide achievement gains or reward them by team when ELL student performance improves.**

One of the points made repeatedly by the advisory group was not to give a false sense of precision in cases where we lack precise measurements. In the case of ELLs, it is probably more appropriate to offer group rewards rather than to reward individual teachers. This is because multiple teachers and instructional staff contribute to an ELL's learning gains, but these gains may not show up for several years. This strategy has the added advantage of emphasizing that the education of ELLs is a responsibility shared by all teachers in the school. Districts might consider rewarding entire teams when academic performance improves, including bilingual and ESL teachers, specialists, and mainstream classroom teachers who work with the school's ELL students. Milanowski reports, for example, that

*"in one of the Teacher Incentive Fund districts, the performance of each elementary grade English language learner is considered the responsibility of 3-5 educators, rather than just the classroom teacher of record. In this situation, the district is working on team level incentives, recognizing that an individual incentive would not reinforce the team concept. Individual incentives would also create a measurement headache when trying to apportion credit for student achievement among team teachers."*⁶⁷

**Option
2****Use student gains in English language proficiency, in addition to gains in subject matter knowledge, as an additional performance measure for teachers of ELLs.**

Though student gains in academic subjects are widely used to identify top-performing teachers, improvements in English language proficiency are less likely to be factored into performance-based teacher pay. Part of the reason for this is that states have only recently been developing statewide tests of English language development. According to the U.S. Government Accountability Office, for example, many states administered new English language development tests for the first time in spring 2006 to meet new assessment requirements specified in the *No Child Left Behind Act*.⁶⁸ During 2005-2006, 14 states reported using an off-the-shelf test; 7 states reported creating their own assessments; 8 states reported working with a test developer to revise an existing test to incorporate their state standards; and 22 reported using a test developed by one of four state consortia. The

advisory group recommended that states and districts consider incorporating these kinds of assessments into performance-based pay formulas to measure the performance of teachers of ELLs.

**Option
3****Use knowledge and skills-based pay structures to reward teachers of English language learners for their expertise.**

An example of a school that has implemented this strategy is the Vaughn Next Century Learning Center, a charter school in Los Angeles that adopted a performance-based pay plan for its teachers in 1998.⁶⁹ The pay plan included a schoolwide performance award that all teachers were eligible to earn, as well as several rewards based on individual teachers' knowledge and skills.

Teachers could earn knowledge and skills-based pay for literacy expertise, ESL or language-development skills, technology skills, special-education inclusion experience, and classroom-management and lesson-planning skills. An administrator, a peer evaluator, and the individual teacher used a four-point scale to assess teacher performance in each area. Ratings were averaged to determine whether the teacher met the performance target and qualified for various knowledge and skills-based salary increases ranging from \$100 to \$2,500. Teachers with specific expertise supporting ELLs were eligible for salary increases of \$1,300 to \$2,500. Those who met minimum thresholds for all areas advanced to a higher tier, where they were eligible for bigger rewards. Teachers who reached the top tier and performed well in all areas could earn up to \$13,100 in additional pay. One potential criticism of this type of reward, of course, is that it is not tied to actual student outcomes.

The following section examines similar measurement challenges and options for rewarding effective teachers of students with disabilities.

TEACHERS OF STUDENTS WITH DISABILITIES

Why is it difficult to measure the effectiveness of teachers of students with disabilities?

There are numerous reasons why it is difficult to measure the effectiveness of teachers of students with disabilities, regardless of whether the teacher is certified as a special education teacher or is a mainstream classroom teacher who happens to have students with disabilities in his or her classes. In the case of a teacher who works exclusively with special education students, the teacher may not have enough tested students to allow a reliable performance measure to be calculated. Another issue for these teachers is that some or all of their students may take an alternate version of the achievement test. In most states, these students make up less than

1 percent of student enrollment. The challenge for school district officials is how to include these students in analyses to determine how much their teachers contributed to student learning.

In the case of regular education teachers who have special education students included in their classes, the issue is that most of these students who receive special education services take district-wide tests, but they do so with accommodations. Whenever accommodations are made to a test, the test is altered, and we cannot assert that the results are on the same scale as the results of those who took the district-wide test without accommodations. This may seem like a fine point, but it is an important one, particularly if the district's compensation system ranks teachers and awards incentive pay only to a set percentage of the top performers. In such a case, teachers who have a lot of accommodated students might be perceived to have an unfair advantage over other teachers.

In both cases, the use of non-traditional instructional arrangements for students with disabilities creates a challenge. The achievement of these students is often the responsibility of a team of teachers, with speech and language therapists, occupational therapists, physical therapists, instructional aides, and others working together on the team.

Finally, we do not know how stable academic achievement gains are from year to year for students with disabilities. As explained in the section on teachers of ELLs, expected gain is based on the assumption that students will perform similarly from year to year. Yet year-to-year learning gains may be highly idiosyncratic for students with disabilities, which is precisely why these students have Individualized Education Programs. As one special education teacher explained:

*"While I wouldn't mind being held accountable for progress, I would have some angst about being held to a year's worth of "normal" growth for an SLD student with a 70 IQ or for a mentally handicapped child. . . If there were some way to define reasonable expectations and measure against the "expected," then that would be fine."*⁷⁰

What are some potential ways to measure the productivity of this group of teachers, and what are the advantages and disadvantages of each of these approaches?

Option 1

Base performance rewards for teachers of students with disabilities on schoolwide achievement gains.

As was the case for ELLs, it is probably more appropriate to offer group rewards to teachers of students with disabilities, rather than to reward individual teachers because multiple teachers and instructional staff

contribute to the learning gains of a student with disabilities. The easiest way to reward groups of teachers of students with disabilities is to compensate all when schoolwide performance increases, in light of the broad range of student ability and the wide variety of individualized education plans, test accommodations, staffing arrangements, and support services provided in special education programs.

**Option
2****Reward teacher teams when the performance of students with disabilities improves.**

Instead of offering bonuses to special education teachers only when the entire school's performance on standardized tests improves, school officials could structure compensation systems so that special education teachers earn team rewards when the performance of students with disabilities improves. Making the target bonus equal is key. Reward systems should not be structured so that special education teachers earn less than other teachers in their schools simply because their particular students are not able to take the standardized tests that the district uses to determine teacher bonuses, as in the following example:

*"I teach Special Ed and worked with preschool students last year who have a variety of disabilities. The students cannot be tracked based on test scores – they are unable to take standardized tests. I will only get money if the scores were raised as a whole in the entire school. I will not get as much money as I deserve just because I happened to teach students that could not take the TAKS and Stanford test."*⁷¹

**Option
3****Develop a new "student sharing" average to assess the performance of special education teachers.**

As members of the advisory group pointed out, special education teachers are rarely in just one classroom. Some special education teachers serve as a resource to mainstream classroom teachers, rather than serve as the teacher-of-record. These teachers may be in and out of the classroom and may not work solely with students who receive special education services.

The idea behind a student sharing average is that it should be possible to identify the proportion of time that special education teachers spend with the various students that are assigned to them. Each student would count as a proportional share. If, for example, the special education teacher spends one-third of his/her time with a particular student, then the classroom effect of that student's teacher-of-record would make up one-third of the special education teacher's value-added score. This type of

student sharing average could also be applied to other support teachers who are helping the whole school. This approach would allow special education teachers to receive credit for the general education students whom they teach when they work in groups, though it would require identification of the percentage of time that the special education teacher spends with each general education teacher's students. It would also address some of the concerns raised by special education teachers whose efforts enable general education teachers to receive awards for raising test scores, yet they receive no bonuses:

*"We work with general education teachers in inclusive settings; however, we did not share in the bonus. . . I know for a fact that without our assistance some of those teachers on my campus would have not been rewarded."*⁷²

*"I do tutor for TAKS and I know of at least one teacher that wouldn't have gotten a large bonus without the Special Education resource teacher helping the students in that class pass TAKS."*⁷³

Option 4

Use knowledge and skills-based pay structures to reward teachers of students with disabilities for their expertise.

As previously noted, the Vaughn Next Century Learning Center in Los Angeles developed an alternative compensation system for teachers that included both a schoolwide performance-pay component and a variety of rewards based on individual teachers' knowledge and skills. One of these knowledge and skills-based rewards was extra pay for special education inclusion experience. Teachers in the lowest tier could earn an additional \$300 if they received a strong evaluation of their expertise in this area. A potential drawback of this option, as mentioned in the discussion of ELLs, is that it does not link teacher compensation to student outcomes.

Measuring the productivity of teachers of students with disabilities is perhaps the single most difficult challenge of designing a performance-pay system that includes all teachers. A limited number of options do exist, but members of the advisory group stressed that states and districts should weigh the merits of more complicated approaches such as the student sharing average against the likelihood that the results would be significantly better, fairer, or more precise than a less complicated approach, such as basing a portion of special educators' pay on schoolwide gains and a portion on principal evaluation. Of the options presented in this section, rewarding teams of teachers when the performance of students with disabilities improves is likely to be the most viable approach.

Summary

As this module has illustrated, the inclusion of the other 69 percent of teachers into performance-pay systems, while daunting, is not impossible. Evidence is clear that the majority of student time is spent in classrooms that are currently not subject to annual testing required by accountability standards. It is important that viable options are provided for measuring the progress of these groups of students and the productivity of their teachers, all of which contribute to school performance.

One way to assess the efficacy of the options that have been presented is the extent to which each suggestion possesses a degree of “policy validity.”⁷⁴ According to Harris, policy validity is the degree to which measures exhibit a high degree of statistical validity, while considering the functions that a measure serves and the costs of producing the measure.

The two primary functions that a measure serves include serving as a signal of those teachers most likely to be effective and as a tool to increase teacher quality by providing formative and summative feedback to teachers. It is important to consider both functions because not every measure that is a good signal of teacher effectiveness provides helpful formative or summative feedback to teachers to facilitate and encourage improvement. This underscores the importance of using multiple measures of student performance and teacher effectiveness in performance-pay plans.

When held against the standard of policy validity, no single strategy that has been described is an ideal solution. However, the strength of a performance-pay system rests substantially upon the use of multiple measures and supports that form a comprehensive, strategic approach to human capital development. Within such a system, each of the suggestions offered in this paper merits consideration as a worthwhile way to include non-core teachers in performance measurement systems. The inclusion of a broader base of teachers in performance-pay systems, by using more appropriate and meaningful measures, endorses several key assumptions of the movement:

1. The performance of all students, including ELLs and students with disabilities, should be assessed and reflected in how teachers are motivated and rewarded.
2. A primary goal of a performance-based compensation system is increased teacher quality and student learning in all grades and subject areas.
3. A reward structure that bases a portion of awards on team or schoolwide growth is likely to encourage collaboration and collective responsibility for all students.
4. Multiple measures and supports should be used to evaluate and encourage teacher performance.
5. Reward structures should not promote adverse effects, such as discouraging teachers from teaching ELLs or students with disabilities.

The table below summarizes the options presented in this paper for including teachers of non-tested subjects and grades, ELLs, and students with disabilities in performance-pay programs. The table makes it apparent that some of the options have broad applicability and are worthwhile for consideration with all four groups of teachers. Others have more specific relevance to individual groups of teachers. What the table provides is an overview of the options available to those charged with designing and implementing performance-pay plans so that they can include the other 69 percent of teachers in the compensation system.

Table 1: Summary of options for including teachers of non-tested subjects and grades, ELLs, and students with disabilities in performance-based compensation systems

Options	Applicable to teachers of:			
	Non-tested subjects	Non-tested grades	English language learners	Students with disabilities
1. Core and non-core teachers are eligible for schoolwide bonuses only	Yes	Yes	Yes	Yes
2. Non-core teachers are eligible for some, but not all, of the individual performance incentives that teachers of core subjects can earn	Yes	Yes	Yes	Yes
3. Student test scores are not used to determine non-core teachers' eligibility for rewards. Instead, eligibility is based exclusively on non-test measures, such as observed evaluations of classroom performance, portfolios, acquisition of additional knowledge and skills, or assumption of additional roles or responsibilities	Yes	Yes	Yes	Yes
4. Create new student tests to assess teacher performance in non-core subjects	Yes	Yes	Yes	Yes
5. Create a developmentally appropriate rubric to assess how well teachers in pre-K to Grade 2 are supporting young children's development on dimensions such as their cognitive development, social-emotional development, motor development, and language acquisition		Yes		
6. Buy off-grade tests from the same vendor that tests Grades 3-8		Yes		
7. Use student results from adaptive tests such as the DIBELS or MAP to assess teacher performance at the early grades		Yes		
8. Base rewards of high school teachers on departmentwide performance, rather than individual classroom performance		Yes		
9. Supplement student test outcomes with schoolwide measures such as high school dropout and graduation rates	Yes	Yes	Yes	Yes
10. Reward teacher teams when the performance of ELLs or students with disabilities improves.			Yes	Yes
11. Use student gains in English language proficiency, in addition to gains in subject matter knowledge, as an additional performance measure for teachers of ELLs.			Yes	
12. Develop a new "student sharing" average to assess the performance of special education teachers				Yes

The incorporation of any of these options will necessitate additional work done by schools or districts to ensure that teachers of non-tested subjects and grades, ELLs, and students with disabilities have the necessary support structures and targeted professional development to meet the measurement goals successfully. In some instances, the options could substantially alter the motivation and reward structure of teachers. In these cases, care must be given to the possibility of goal displacement and of unintended consequences associated with additional measures of teacher performance. In all cases, states and districts should consider the degree to which the measures used in a performance-pay plan:

- provide indicators of teacher effectiveness and provide teachers with feedback to encourage their ongoing development;
- are part of an integrated system designed to support and develop human capital; and
- are clearly defined and aligned with goals for effective instruction and student learning across the curriculum.

We conclude by offering the following recommendations to help states, districts, and schools consider how to include the other 69 percent of teachers in performance-pay plans:

- **Consider fairness from multiple perspectives.** All teachers must have an opportunity to earn awards for their contribution to student performance, regardless of the grade, subject, and types of students they teach. However, it is not necessary that access to rewards be even.
- **Think systemically and holistically.** It is important that ongoing support and professional development are provided to help teachers meet growth targets and to ensure that growth targets are directly linked to school and district goals.
- **Include multiple measures and award types.** Consider the advantages of using multiple measures of student performance and teacher effectiveness and the benefits of awards at the individual, team, department, and school levels.
- **Weigh transparency versus accuracy.** States and districts need to make difficult decisions regarding the inclusion of measures that cannot be readily explained to key constituents or that cannot be verified as valid and reliable measures of teacher effectiveness. These trade-offs often hinge on the degree to which measures exhibit a necessary level of technical rigor while also remaining understandable to those whom they directly affect.

- **Be aware of potential unintended consequences.** Within a performance-pay system, the manner in which teacher effectiveness and student performance are measured will influence teacher motivation. States and districts should consider how the system affects teachers at all grade levels, subject areas, and across the spectrum of student ability levels and demographic characteristics.

End Notes

- ¹ Crouse, J. "Points would decide bonuses," *The Ledger*, April 6, 2006.
- ² Alaska Department of Education and Early Development. (2006, January 11). *Alaska School Performance Incentive Program: Measuring individual achievement*. PowerPoint presentation.
- ³ TeacherSolutions. (2007). *Performance-pay for teachers: Designing a system that students deserve*. Hillsborough, NC: Center for Teaching Quality. <http://www.teachingquality.org/pdfs/TSreport.pdf>
- ⁴ Studies of student achievement in mathematics: Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1); Studies of student achievement in mathematics and reading/language arts: Muralidaran, K., & Sundararaman, V. (2006, November). *Teacher incentives in developing countries: Experimental evidence from India*. Cambridge, MA: Harvard University, Department of Economics; Barnett, J.H., Ritter, G.W., Winters, M.A., & Greene, J.P. (2007, January). *Evaluation of year one of the achievement challenge pilot project in the Little Rock Public School District*. Fayetteville, AR: University of Arkansas, Department of Education Reform. http://www.uark.edu/ua/der/Research/merit_pay.html; Boozer, M.A. (1999). *The design and evaluation of incentive schemes for schools: Evidence from South Carolina's Teacher Incentive Pay Project*. Mimeo, Hoover Institute; Clotfelter, C., & Ladd, H. (1996). Recognizing and rewarding success in public schools. In H.F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education*, pp. 23-64. Washington, DC: The Brookings Institution; Cooper, S.T., & Cohn, E. (1997, June). Estimation of a frontier production function for the South Carolina educational process. *Economics of Education Review*, 16(3), 313-327; Dee, T.S., & Keys, B.J. (2005). Dollars and sense: What a Tennessee experiment tells us about merit pay. *Education Next*, 5(1) 60-67. http://media.hoover.org/documents/ednext20051_60.pdf; Koedel, C. (2007, September). *Teacher quality and educational production in secondary school*. Working Paper 2007-2. Nashville, TN: Vanderbilt University, National Center on Performance Incentives. http://www.performanceincentives.org/data/files/news/PapersNews/Koedel_2007a_Revised.pdf; Ladd, H.F. (1999, February). The Dallas school accountability and incentive program: An evaluation of its impacts on student outcomes. *Economics of Education Review*, 18(1), 1-16; Ritter, G., Holley, M., Jensen, N., Riffel, B., Winters, M., Barnett, J., & Greene, J. (2008, January). Year two evaluation of the achievement challenge pilot project in the Little Rock Public School District. Fayetteville, AR: University of Arkansas, Department of Education Reform. http://www.uark.edu/ua/der/Research/merit_pay/year_two.html; Winters, M.A., Ritter, G.W., Barnett, J.H., & Greene, J.P. (2007, January). *An evaluation of teacher performance pay in Arkansas*. Fayetteville, AR: University of Arkansas, Department of Education Reform. http://www.uark.edu/ua/der/Research/performance_pay_ar.html
- ⁵ Studies of student achievement in science and history, as well as mathematics and reading/language arts: Atkinson, A., Burgess, S., Croxon, B., Gregg, P., Propper, C., Slater, H., & Wilson, D. (2004, December). *Evaluating the impact of performance-related pay for teachers in England*. Working Paper No. 04/113. University of Bristol: Centre for Market and Public Organization; Figlio, D.N., & Kenny, L.W. (2007, June). Individual teacher incentives and student performance. *Journal of Public Economics*, 91(5-6), 901-914.
- ⁶ Ballou, D. (2002, Summer). Sizing up test scores. *Education Next*, 2(2). <http://www.hoover.org/publications/ednext/3365706.html>
- ⁷ Koedel, C. (2007, September). *Teacher quality and educational production in secondary school*. Working Paper 2007-2. Nashville, TN: Vanderbilt University, National Center on Performance Incentives. http://www.performanceincentives.org/data/files/news/PapersNews/Koedel_2007a_Revised.pdf
- ⁸ *Ibid.*
- ⁹ *Ibid.*

- ¹⁰ Chait, R. (2007, November). *Current state policies that reform teacher pay: An examination of pay-for-performance programs in eight states*. Washington, DC: Center for American Progress.
http://www.americanprogress.org/issues/2007/11/pdf/teacher_pay.pdf
- ¹¹ Goldhaber, D. (2006, December). *Teacher pay reforms: The political implications of recent research*. Washington, DC: Center for American Progress. p. 19.
http://www.americanprogress.org/issues/2006/12/pdf/teacher_pay_report.pdf
- ¹² See Holmstrom, B., & Milgrom, P. (1991). Multi-task principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization*, 7, 24-52.
- ¹³ Koretz, D.M. (2002, Autumn). Limitations in the use of achievement tests as measures of educators' productivity. *Journal of Human Resources*, 37(4), 752-777. See for example Darling-Hammond, L., & Wise, A.E. (1985). Beyond standardization: State standards and school improvement. *The Elementary School Journal*, 85(3), 315-336; Koretz, D.M., Barron, S., Mitchell, K., & Stecher, B. (1996a). *The perceived effects of the Kentucky Instructional Results Information System (KIRIS)*. MR-792-PCT/FF. Santa Monica: RAND; Koretz, D.M., Mitchell, K., Barron, S., & Keith, S. (1996b). *The perceived effects of the Maryland School Performance Assessment Program. CSE Technical Report No. 409*. Los Angeles: University of California, Center for the Study of Evaluation; Salmon-Cox, L. (1982). *MAP Math: End of year one report*. University of Pittsburgh, Learning Research and Development Center, Unpublished; Salmon-Cox, L. (1984). *MAP Reading End-of-Year report*. University of Pittsburgh, Learning Research and Development Center, Unpublished; Shepard, L.A. (1988, April). *The harm of measurement-driven instruction*. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC; Shepard, L.A., & Dougherty, K.D. (1991, April). Effects of high-stakes testing on instruction. In R.L. Linn (Chair), *The effects of high-stakes testing*. Symposium presented at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago; Stecher, B.M., & Barron, S.I. (1999). *Quadrennial milepost accountability testing in Kentucky*. CSE Technical Report No. 505. Los Angeles: University of California, Center for the Study of Evaluation.
- ¹⁴ TeacherSolutions. (2007). *Performance-pay for teachers: Designing a system that students deserve*. Hillsborough, NC: Center for Teaching Quality. p. 40.
<http://www.teachingquality.org/pdfs/TSreport.pdf>
- ¹⁵ School Zone. "And the winners are..." *Houston Chronicle* Blog, January 23, 2007.
- ¹⁶ Chait, R. (2007, November). *Current state policies that reform teacher pay: An examination of pay-for-performance programs in eight states*. Washington, DC: Center for American Progress.
http://www.americanprogress.org/issues/2007/11/pdf/teacher_pay.pdf
- ¹⁷ See Milanowski, A. (2008). *How to pay teachers for student performance outcomes*. Madison, WI: Consortium for Policy Research in Education (CPRE), University of Wisconsin-Madison.
- ¹⁸ Koedel, C. (2007, September). *Teacher quality and educational production in secondary school*. Working Paper 2007-2. Nashville, TN: Vanderbilt University, National Center on Performance Incentives.
http://www.performanceincentives.org/data/files/news/PapersNews/Koedel_2007a_Revised.pdf
- ¹⁹ Houston Independent School District. "Teacher-performance-pay program receives enthusiastic Board approval." Press release, January 13, 2006.
- ²⁰ School Zone. "And the winners are..." *Houston Chronicle* Blog, January 23, 2007.
- ²¹ *Ibid.*
- ²² *Ibid.*
- ²³ Comments made by Carla Stevens, Houston Independent School District, at the U.S. Department of Education's Teacher Incentive Fund grantee meeting, June 5, 2008.

- ²⁴ See Milanowski, A. (2008). *How to pay teachers for student performance outcomes*. Madison, WI: Consortium for Policy Research in Education (CPRE), University of Wisconsin-Madison.
- ²⁵ See Guilford County Schools. "Mission Possible pay incentives." http://www.gcsnc.com/depts/mission_possible/pay.htm
- ²⁶ Comments made by Emily Scott and Holly Bayonas, Guilford County Schools, at the U.S. Department of Education's Teacher Incentive Fund grantee meeting, June 5, 2008.
- ²⁷ Winnick, P. "Teachers unhappy tying pay to results," *Pittsburgh Post-Gazette*, August 31, 2000. <http://www.post-gazette.com/regionstate/20000831bonus3.asp>
- ²⁸ LaFee, S. (2000, October). Linking teacher pay to student scores. *The School Administrator*. <http://www.aasa.org/publications/saarticledetail.cfm?ItemNumber=4158>
- ²⁹ Winnick, P. "Teachers unhappy tying pay to results," *Pittsburgh Post-Gazette*, August 31, 2000. <http://www.post-gazette.com/regionstate/20000831bonus3.asp>
- ³⁰ See Jacob, B.A., & Lefgren, L. (2008, January). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(1), 101-136.
- ³¹ See Halverson, R., Kelley, C., & Kimball, S. (2004). Implementing teacher evaluation systems: How principals make sense of complex artifacts to shape local instructional practice. In W. Hoy & C. Miskel (Eds.), *Research and theory in educational administration*. (Vol. 3). Greenwich, CT: Information Age Publishing; Nelson, B.S., & Sassi, A. (2000). Shifting approaches to supervision: The case of mathematics supervision. *Educational Administration Quarterly*, 36(4), 553-584; Peterson, K.D. (2000). *Teacher evaluation: A comprehensive guide to new directions and practice*. (2nd ed.). Thousand Oaks, CA: Corwin Press. See also Keung Hui, T. "Easy A's for Wake teachers?" *The (Raleigh, NC) News & Observer*, September 6, 2007; The New Teacher Project. (2007, July). *Teacher hiring, transfer and assignment in Chicago Public Schools*. New York: Author. <http://www.tntp.org/ourresearch/documents/TNTPAnalysis-Chicago.pdf>; Rossi, R. "Teachers' grades inflated," *Chicago Sun-Times*, July 30, 2007. <http://www.sun-times.com/>
- ³² TeacherSolutions. (2007). *Performance-pay for teachers: Designing a system that students deserve*. Hillsborough, NC: Center for Teaching Quality. <http://www.teachingquality.org/pdfs/TSreport.pdf>
- ³³ See Milanowski, A. (2008). *How to pay teachers for student performance outcomes*. Madison, WI: Consortium for Policy Research in Education (CPRE), University of Wisconsin-Madison.
- ³⁴ For a case summary of performance pay in Florida, see Max, J. (2007, November). *The evolution of performance pay in Florida*. Washington, DC: Center for Educator Compensation Reform. <http://www.cccr.ed.gov/guides/summaries/FloridaCaseSummary.pdf>
- ³⁵ Whoriskey, P. "Fla. to link teacher pay to students' test scores; Critics worry about fairness," *Washington Post*, March 22, 2006.
- ³⁶ Pinzur, M. "State education board approves merit-pay plan for teachers," *Miami Herald*, February 21, 2006.
- ³⁷ Pinzur, M. "State education board approves merit-pay plan for teachers," *Miami Herald*, February 21, 2006; Florida School Boards Association. (2006, March). Teachers appeal performance pay. *Leaders*, 7(3), 1, 3.
- ³⁸ Lewis, K. "Collier teachers slam bonus plan based on FCAT," *Naples News*, February 11, 2006; Winchester, D. "FCAT bonus plan for teachers wins approval," *St. Petersburg Times*, February 22, 2006; Pinzur, M. "Performance-based pay for teachers OK'd," *Miami Herald*, February 22, 2006; Crouse, J. "Some teachers say merit pay plan is a bitter apple," *The Ledger*, March 15, 2006.

- ³⁹ Rodriguez, E. "Panel members to discuss teachers' pay plan today," *Bradenton Herald*, April 6, 2006.
- ⁴⁰ Pinzur, M. "State education board approves merit-pay plan for teachers," *Miami Herald*, February 21, 2006; Pinzur, M. "Performance-based pay for teachers OK'd," *Miami Herald*, February 22, 2006.
- ⁴¹ Sampson, H. "State links some teacher bonuses to final exams," *Miami Herald*, November 29, 2006.
- ⁴² Comments made by Joe Perez, Hillsborough County Public Schools, at the U.S. Department of Education's Teacher Incentive Fund grantee meeting, June 5, 2008.
- ⁴³ Solochek, J.S. "Teachers' pay plan revised," *St. Petersburg Times*, September 1, 2007.
- ⁴⁴ Sampson, H. "State links some teacher bonuses to final exams," *Miami Herald*, November 29, 2006.
- ⁴⁵ Florida Department of Education. (n.d.). *2006-07 STAR program: Using FCAT Reading as a pre-test for non-FCAT subjects in a value table*.
<http://www.fldoe.org/PerformancePay/pdfs/FCATReadingPretest.pdf>
- ⁴⁶ See Max, J. (2007, November). *The evolution of performance pay in Florida*. Washington, DC: Center for Educator Compensation Reform.
<http://www.cecr.ed.gov/guides/summaries/FloridaCaseSummary.pdf>
- ⁴⁷ Hobbs, E. "German exam in French is all Greek to kids," *Orlando Sentinel*, May 4, 2007; Hobbs, E. "Teachers slam state merit-pay program," *Orlando Sentinel*, August 20, 2007; Hobbs, E. "Teachers of electives don't see many bonuses," *Orlando Sentinel*, September 9, 2007; Hobbs, E. "Which teachers should win a bonus?" *Sentinel School Zone*, *Orlando Sentinel*, September 10, 2007.
- ⁴⁸ Marshall, T. "Schools may drop merit pay," *St. Petersburg Times*, November 5, 2007.
http://www.sptimes.com/2007/11/05/Hernando/Schools_may_drop_merit.shtml;
Marshall, T. "Merit pay for Hernando teachers heads to new vote," *St. Petersburg Times*, November 7, 2007.
http://www.sptimes.com/2007/11/07/Hernando/Merit_pay_for_Hernand.shtml;
Marshall, T. "Teachers, district cheer: Merit pay plan quashed," *St. Petersburg Times*, November 30, 2007.
http://www.sptimes.com/2007/11/30/Hernando/Teachers_district_ch.shtml
- ⁴⁹ Weber, D. "Merit pay bypasses area teachers," *Orlando Sentinel*, January 11, 2008.
http://www.orlandosentinel.com/news/education/orl-meritpay1108jan11_0,5638597.story
- ⁵⁰ See Max, J. (2007, November). *The evolution of performance pay in Florida*. Washington, DC: Center for Educator Compensation Reform.
<http://www.cecr.ed.gov/guides/summaries/FloridaCaseSummary.pdf>
- ⁵¹ See for example Shepard, L.A., Kagan, S.L., & Wurtz, E.O. (1998). *Principles and recommendations for early childhood assessments*. Washington, DC: National Education Goals Panel.
- ⁵² Clotfelter, C.T., Ladd, H.F., & Vigdor, J.L. (2007, November). *Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects*. NBER Working Paper 13617. Cambridge, MA: National Bureau of Economic Research.
<http://www.nber.org/papers/w13617>
- ⁵³ Koedel, C. (2007, September). *Teacher quality and educational production in secondary school*. Working Paper 2007-2. Nashville, TN: Vanderbilt University, National Center on Performance Incentives. http://www.performanceincentives.org/data/files/news/PapersNews/Koedel_2007a_Revised.pdf
- ⁵⁴ Koretz, D.M. (2002, Autumn). Limitations in the use of achievement tests as measures of educators' productivity. *Journal of Human Resources*, 37(4), 752-777.
- ⁵⁵ Winnick, P. "Teachers unhappy tying pay to results," *Pittsburgh Post-Gazette*, August 31, 2000. <http://www.post-gazette.com/regionstate/20000831bonus3.asp>

- ⁵⁶ La Paro, K.M., Pianta, R.C., & Stuhlman, M. (2004). The classroom assessment scoring system: Findings from the pre-kindergarten year. *The Elementary School Journal*, 104(5), 409-426.
- ⁵⁷ Glover, M.J. "76 educators receive incentive bonuses," *Greensboro News-Record*, October 30, 2007. <http://www.news-record.com/apps/pbcs.dll/article?AID=/20071030/NRSTAFF/71029033>
- ⁵⁸ See Guilford County Schools. "Mission Possible pay incentives." http://www.gcsnc.com/depts/mission_possible/pay.htm
- ⁵⁹ Guilford County Schools. "Mission Possible K-2 Regular Classroom Teacher Agreement." [http://www.gcsnc.com/depts/mission_possible/pdf/MP_Contract_C - ES K-2.pdf](http://www.gcsnc.com/depts/mission_possible/pdf/MP_Contract_C_-_ES_K-2.pdf)
- ⁶⁰ Northwest Evaluation Association (NWEA). "Measures of Academic Progress (MAP): Frequently Asked Questions." <http://www.nwea.org/assessments/map.asp>
- ⁶¹ Comments made by Carla Stevens, Houston Independent School District, at the U.S. Department of Education's Teacher Incentive Fund grantee meeting, June 5, 2008.
- ⁶² Doran, H.C., & Izumi, L.T. (2004, June). *Putting education to the test: A value-added model for California*. San Francisco: Pacific Research Institute.
- ⁶³ Hakuta, K., Butler, Y.G. & Witt, (2000), cited in Gandara, P., Rumberger, R., Maxwell-Jolly, J., & Callahan, R. (2003, October 7). English learners in California schools: Unequal resources, unequal outcomes. *Education Policy Analysis Archives*, 11(36). <http://epaa.asu.edu/epaa/v11n36/>
- ⁶⁴ Thompson, M.S., DiCerbo, K.E., Mahoney, K. & MacSwan, J. (2002, January 25). ¿*Exito en California?* A validity critique of language program evaluations and analysis of English learner test scores. *Education Policy Analysis Archives*, 10(7). <http://epaa.asu.edu/epaa/v10n7/>
- ⁶⁵ Resnick, L.B., & Resnick, D.P. (1990). Tests as standards of achievement in schools. In G. R. Anrig (Ed.), *The uses of standardized tests in American education: Proceedings of the 1989 ETS Invitational Conference* (pp. 6380). Princeton, NJ: Educational Testing Service.
- ⁶⁶ Gandara, P., & Merino, B. (1993). Measuring the outcomes of LEP programs: Test scores, exit rates, and other mythological data. *Educational Evaluation and Policy Analysis*. 15(3): 320-338.
- ⁶⁷ Milanowski, A. (2008). *How to pay teachers for student performance outcomes*. Madison, WI: Consortium for Policy Research in Education (CPRE), University of Wisconsin-Madison.
- ⁶⁸ Ashby, C.M. (2007, March 23). *No Child Left Behind Act: Education assistance could help states better measure progress of students with limited English proficiency*. Testimony before the Subcommittee on Early Childhood, Elementary and Secondary Education, Committee on Education and Labor, House of Representatives. GAO-07-646T. <http://www.gao.gov/new.items/d07646t.pdf>
- ⁶⁹ Odden, A. (2001, Spring). Rewarding expertise. *Education Next*, 1(1). <http://www.hoover.org/publications/ednext/3390871.html>
- ⁷⁰ Teacher Leaders Network. "TLN Voices: Exploring teacher compensation and pay for performance."
- ⁷¹ School Zone. "And the winners are..." *Houston Chronicle* Blog, January 23, 2007.
- ⁷² *Ibid.*
- ⁷³ *Ibid.*
- ⁷⁴ The concept of policy validity is introduced and explained in detail by Douglas N. Harris in his 2008 manuscript titled The Policy Uses and "Policy Validity" of Value-Added and Other Teacher Quality Measures. Forthcoming in D.H. Gitomer (Ed.), *Measurement Issues and the Assessment of Teacher Quality*. Thousand Oaks, CA: SAGE Publications.